INTRODUCTION TO ARTIFICIAL INTELLIGENCE (BDM 3014)

PROJECT PROPOSAL

PROJECT TITLE:

CREDIT CARD FRAUD DETECTION

PROFESSOR: BHAVIK GANDHI

GROUP - 1

AYUSH JOSHI (C0905800)

HARSIMRAN KAUR (C0908419)

RUTVIK PRAJAPATI (C0909130)

SALONI SHAH (C0901921)

SRIVENI KUNDURU (C0876299)

TAMANNA (C0909095)

TEGHBEER SINGH (C0901923)

TABLE OF CONTENT

1. ABOUT THE INDUSTRY

Credit Card Fraud Industry

In the complex world of credit card fraud, offenders use a variety of advanced strategies to take advantage of weaknesses in the financial system. Every stage of the fraud process, from gathering cardholder data to carrying out fraudulent transactions, is carefully planned to minimize detection and maximize illegal profits. Here, we explore the intricate details of the credit card fraud sector, illuminating its inner workings and the difficulties it presents to cardholders and financial institutions across the globe.

- **Obtaining Cardholder Information:** Sophisticated methods are employed by fraudsters to get cardholder information. They might put in skimming devices on ATMs or point-of-sale terminals, send phishing emails or messages to people pretending to be the owner of a card, or take advantage of system flaws to access databases that hold cardholder data without authorization (data breaches).
- **Card Cloning and Fabrication:** After obtaining cardholder data, fraudsters use the stolen data to either clone or construct actual cards. This procedure entails either making counterfeit cards from scratch or encoding the details from the stolen card onto blank cards, frequently utilizing advanced tools and methods.
- **Fraudulent purchases:** Using stolen or counterfeit credit cards, con artists conduct illicit physical and online purchases. They may use fraudulent cards to make purchases at physical places, buy online, or take out cash from ATMs, frequently resulting in large losses for cardholders as well as financial institutions.
- **Anonymization strategies:** Fraudsters use a variety of anonymization strategies to avoid discovery and hide their identity. These could be carrying out transactions with anonymizing cryptocurrencies like Bitcoin, utilizing virtual private networks (VPNs) to conceal their IP addresses, or directing transactions through proxy servers to hide their whereabouts.
- **Card Not Present (CNP) Fraud:** In online transactions, thieves use card information that has been stolen to make purchases without the card being physically present. They employ strategies include account takeover, in which they breach victims' internet accounts to carry out fraudulent transactions covertly, and card testing, in which information that has been hacked is utilized to confirm purchases.
- **Money Laundering:** Fraudsters use intricate money laundering strategies to hide their illicit gains and launder the proceeds of credit card fraud. This could entail making numerous account transfers, establishing shell corporations to hide the money's true source, or making a lot of transactions to hide the trail of money laundering.
- **Collusion and Insider Fraud:** Sometimes, people working for banks or other organizations would band together with outside scammers to carry out fraudulent activities. Employees who engage in insider fraud use their access to and familiarity with internal systems to commit crimes, which makes it difficult to detect and stop fraud.

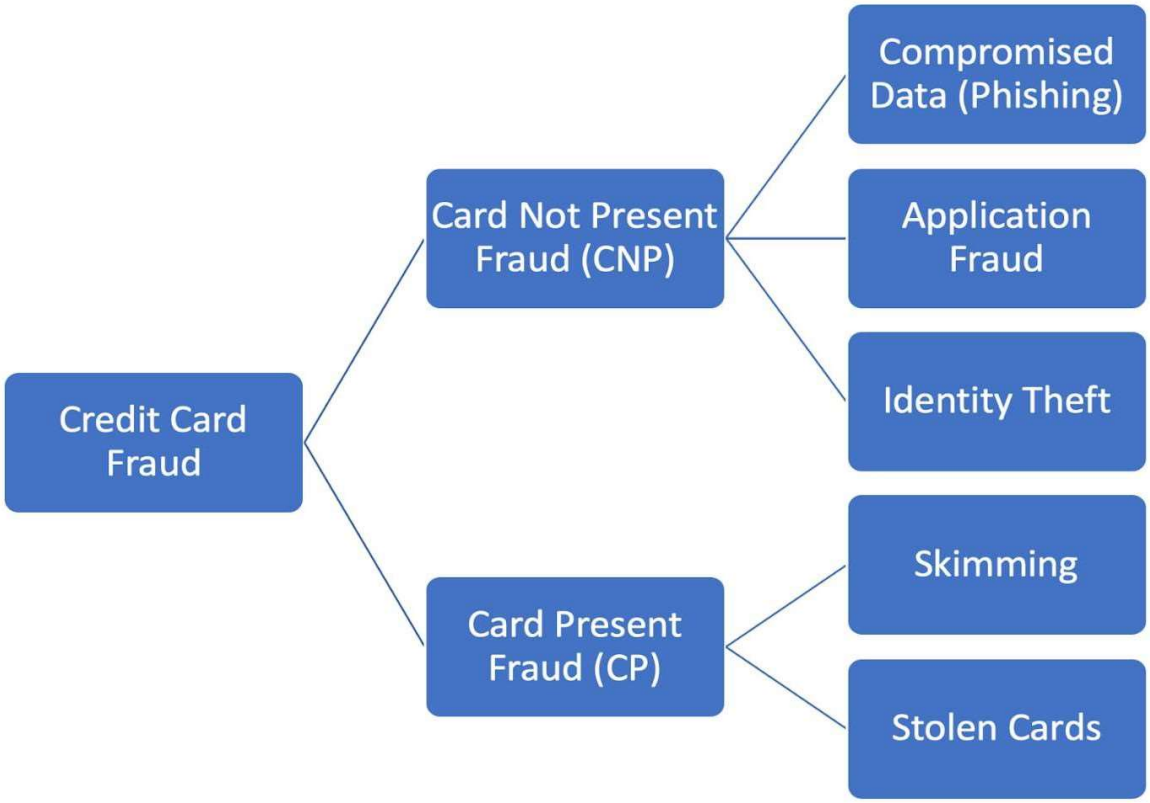

Figure 1: Different methods used for credit card fraud

The credit card fraud detection industry

- **Data Collection:** Credit card transactions, user profiles, past fraud records, and other sources provide financial institutions with a wealth of transactional data, including cardholder information, transaction history, and behavioral trends.
- **Preprocessing of the Data:** The gathered data is cleaned, normalized, and formatted to create a structured dataset that is ready for analysis. To guarantee correctness and consistency, this entails eliminating discrepancies, managing missing information, and standardizing data formats.
- **Using feature engineering:** The pertinent information that can differentiate between authentic and fraudulent transactions is extracted from the preprocessed data. These characteristics, which offer useful inputs for fraud detection models, may include transaction amounts, frequency, temporal fluctuations, and user behavior patterns.
- **Model Training and Evaluation:** To identify patterns and anomalies suggestive of fraudulent activity, sophisticated machine learning algorithms—such as logistic regression, decision trees, random forests, and neural networks—are trained using transaction data from previous transactions. Metrics like accuracy, precision, recall, and F1 score are used to analyze these models to determine how well they perform and identify fraud.
- **Real-Time Monitoring:** Fraud detection models that have been deployed keep an eye on incoming transactions constantly, assigning a score according to how likely it is that they are involved in fraud. Alerts for additional inquiry are set off by suspicious activity, enabling financial institutions to move quickly to reduce fraud losses and safeguard cardholders.

- **Updating and Maintaining the Model:** To stay successful and adjust to changing fraud patterns, fraud detection models need constant upkeep. To increase accuracy and efficiency, financial institutions track model performance regularly, consider input from fraud cases they have identified, and change their models as needed.
- **Cooperation and Information Sharing:** To exchange fraud intelligence and fight fraud collectively, financial institutions work with law enforcement agencies, industry stakeholders, and regulatory bodies. Initiatives for information sharing make it easier to share fraud trends, best practices, and practical insights, which improves the efficiency of fraud detection activities in the sector.

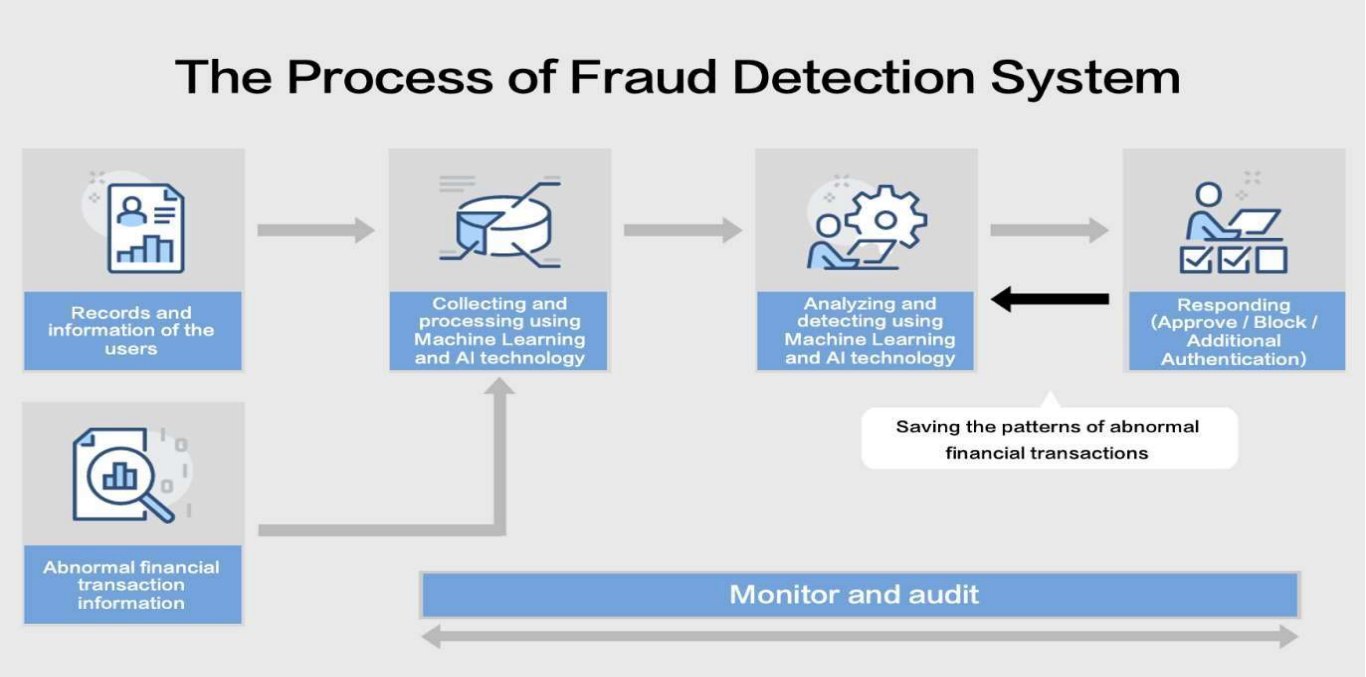

Figure 2: Step by step process for Fraud Detection

2. PROBLEM STATEMENT

Problem Overview:

Accurately and swiftly differentiating between authentic and fraudulent transactions is the main problem in credit card fraud detection. To commit fraud, scammers use a variety of strategies, including identity theft, card-not-present transactions, and stolen card details. Sophisticated fraud schemes are frequently too complex for conventional rule-based systems and simple anomaly detection techniques to identify. Furthermore, fraud detection systems need to adjust to latest trends and patterns in fraudulent behaviour as con artists refine their methods.

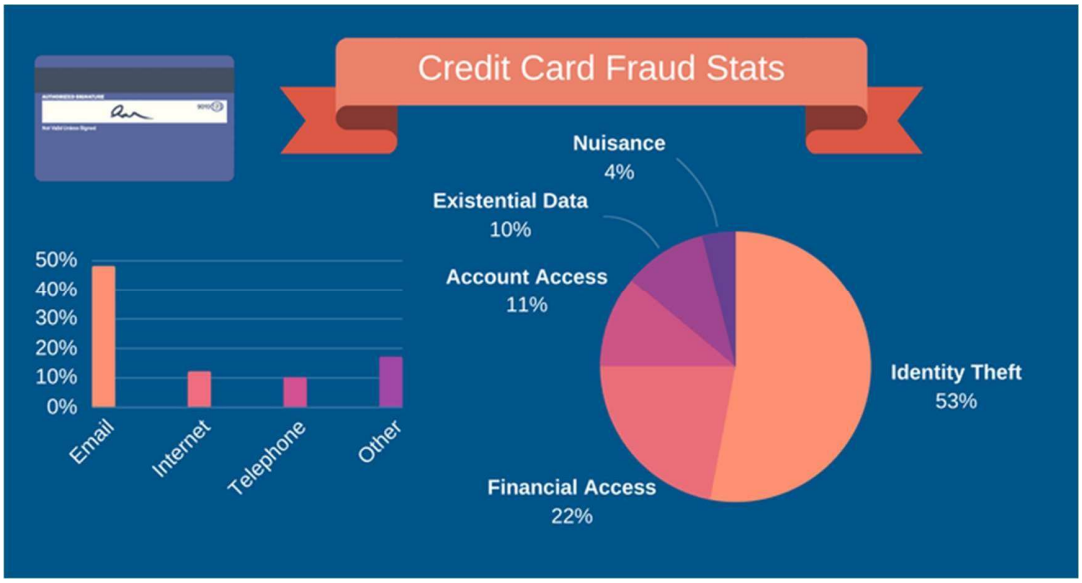

Figure 3: Statistics of different strategies used for committing credit card fraud

Current Limitations:

Existing fraud detection systems encounter several limitations that hinder their effectiveness. These limitations include:

1. **False Positives:** Existing systems frequently produce false positives, marking legal transactions as fraudulent, which causes inconvenience to the customers and raises operating expenses for companies.
2. **Delayed Detection:** Some fraud detection systems have latency problems, which cause fraudulent transactions to go undetected for longer periods of time and provide fraudsters the opportunity to carry out several illegal activities.
3. **Lack of Scalability:** When transaction volume rises, traditional fraud detection techniques find it difficult to grow, which results in decreased accuracy and performance degradation.
4. **Complex Fraud Patterns**: Since fraudsters are always changing their strategies, it is difficult for traditional systems to stay up to date with the latest trends and patterns in fraud.

Proposed Solution:

To overcome the flaws of existing fraud detection systems, this project suggests creating an enhanced credit card fraud detection system to address the flaws of the current systems, **specifically for Europe as we are working with dataset from Europe for the year 2013.** The system will examine transactional data and find trends suggestive of fraudulent activity by utilising innovative machine learning methods, including supervised and unsupervised learning approaches like random forest, logistic regression, and decision tree regression. The system's ability to monitor in real-time will help it identify suspicious activity quickly, reducing the possibility of financial losses.

Objectives:

The objectives of this project are as follows:

1. Developing several machine learning models that can effectively identify fraudulent credit card transactions while reducing the number of false positives.
2. Putting in place alerting and monitoring systems that operate in real-time to quickly spot and prevent such fraud attempts.
3. Boost the fraud detection system's efficiency and scalability to manage high transaction volumes without compromising speed.
4. Update the system often to accommodate evolving fraud trends and scammers' changing strategies.

3. DIFFERENTIATION

In a crowded arena of initiatives, our project distinguishes itself through its unparalleled blend of vision, execution, and transformative potential in following ways:

- **Utilization of Advanced Technologies:** Our Credit card fraud detection project leverage innovative technologies such as artificial intelligence (AI), machine learning (ML), and data analytics. These technologies enable the development of sophisticated algorithms capable of identifying complex patterns and anomalies indicative of fraudulent activity, setting it apart from traditional rule-based systems.
- **User-Friendly Interface and Transparency:** This project prioritize user experience by offering intuitive and user-friendly interfaces for end-users. It provides transparent insights into the fraud detection process, allowing stakeholders to understand how decisions are made and fostering trust in the system's capabilities.

4. DATA GATHERING

- The dataset contains credit card transactions made by European cardholders in the year 2013.
- It comprises over 284,807 records, providing a substantial amount of data for analysis.
- The data has been anonymized to protect the cardholders' identities, ensuring compliance with privacy regulations.

Data Look:

- The dataset is structured in tabular form, with rows representing individual transactions and columns representing unique features.
- Each row corresponds to a single credit card transaction, where **time** contains the seconds elapsed between each transaction and the first transaction in the dataset.
- The **V1-V28** has anonymized features representing various transaction attributes. These features include details such as time, location, and other relevant transaction characteristics.
- The **Amount** column provides numerical data representing the monetary value of each transaction, which is crucial for fraud detection algorithms.

- The **Class** column serves as the target variable for machine learning models, indicating whether a transaction is fraudulent or not.

```
Dataset Information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 284807 entries, 0 to 284806
Data columns (total 31 columns):
 #   Column  Non-Null Count   Dtype
---  ------  --------------   -----
 0   Time    284807 non-null  float64
 1   V1      284807 non-null  float64
 2   V2      284807 non-null  float64
 3   V3      284807 non-null  float64
 4   V4      284807 non-null  float64
 5   V5      284807 non-null  float64
 6   V6      284807 non-null  float64
 7   V7      284807 non-null  float64
 8   V8      284807 non-null  float64
 9   V9      284807 non-null  float64
 10  V10     284807 non-null  float64
 11  V11     284807 non-null  float64
 12  V12     284807 non-null  float64
 13  V13     284807 non-null  float64
 14  V14     284807 non-null  float64
 15  V15     284807 non-null  float64
 16  V16     284807 non-null  float64
 17  V17     284807 non-null  float64
 18  V18     284807 non-null  float64
 19  V19     284807 non-null  float64
 20  V20     284807 non-null  float64
 21  V21     284807 non-null  float64
 22  V22     284807 non-null  float64
```

```
df.isnull().values.any()

False
```

Through this, we can see that there are no null values present in our dataset.

Summary:

- The dataset provides a comprehensive overview of credit card transactions, allowing for in-depth analysis and modeling to detect and prevent fraudulent activities.
- With its enormous size and anonymized features, the dataset presents ample opportunities for exploring various machine learning algorithms and techniques for credit card fraud detection.
- Additionally, the dataset's structure facilitates potential analyses on merchant categories and transaction types to uncover patterns associated with fraud.

5. WIREFRAME OF THE PROJECT

a. Data Preprocessing:

- Data cleaning steps such as handling missing values, outlier detection, and normalization.
- Feature engineering techniques to extract additional relevant features if necessary.

b. Model Selection:

- Section displaying different machine learning models considered for the project (e.g., logistic regression, random forest, support vector machines, neural networks).
- Brief description of each model's strengths and weaknesses.

c. Model Training:

- Training each selected model using the training dataset.
- Hyperparameter tuning if applicable, using techniques like grid search or random search.

d. Model Evaluation:

- Evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC score for each model.
- Visualizations like ROC curves or precision-recall curves for model comparison.

e. Feature Importance:

- Visualization of feature important scores for models that support it (e.g., random forest).
- Highlighting the most key features for fraud detection.

f. Model Deployment:

- Deploying the models using Flask or using cloud-based services like Sagemaker etc.

g. Performance Analysis:

- Evaluation of the model's performance on the test dataset, including any insights gained from analyzing misclassifications.

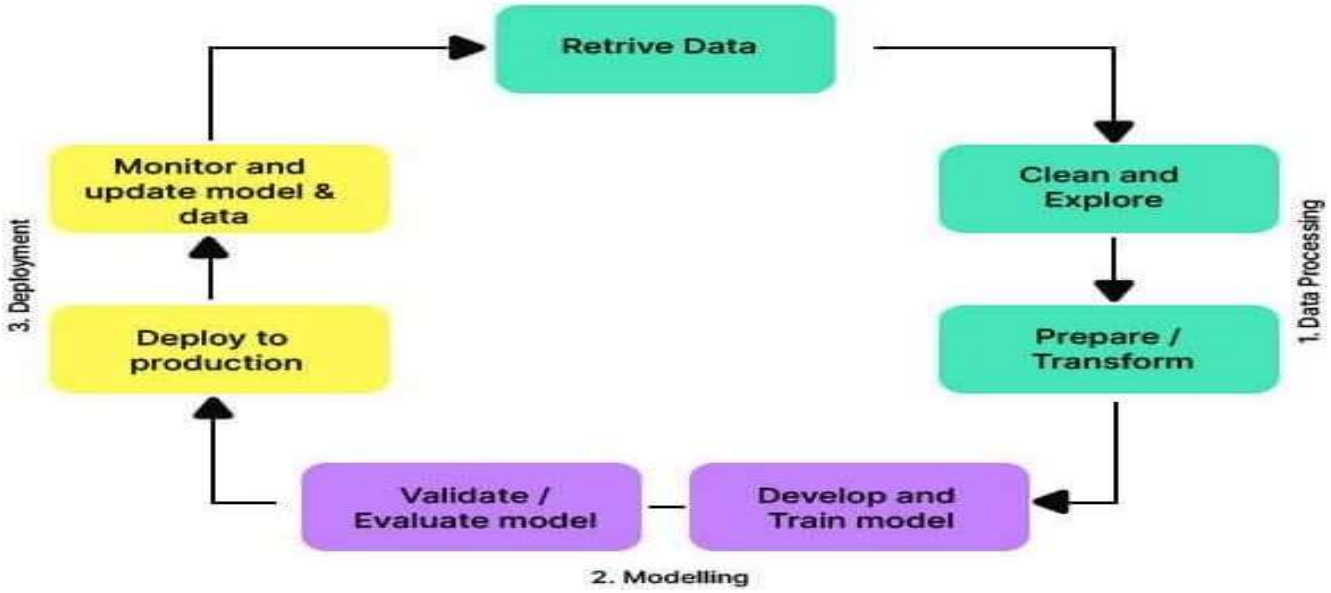

Figure 4: The Workflow of the Project

User Interface Workflow:

a. **Input Interface: The** UI will provide an interface for users to input the necessary information for fraud detection. This typically includes the transaction attributes such as time, transaction amount, and other anonymized features (V1-V28).

b. **Model Selection: The** UI would allow users to choose which machine learning model they want to use for fraud detection. This can include options such as Logistic Regression, Random Forest, Support Vector Machines (SVM), Gradient Boosting, etc. Users may select the model based on their preference or the model's performance.

c. **Data Preprocessing**: Once the user inputs the transaction attributes, the UI would preprocess this data to make it suitable for input into the selected machine learning model. This may involve handling missing values, scaling, or normalizing features, and any other necessary data transformations.

d. **Model Prediction**: After preprocessing the data, the UI should use the selected machine learning model to make predictions on whether the transaction is fraudulent or not. The model will take the preprocessed input data as input and output a binary label indicating the likelihood of fraud.

e. **Result Display**: The UI will display the results of the fraud detection process to the user. This can include the model's prediction (fraudulent or not fraudulent), as well as any additional information such as the probability or confidence score associated with the prediction.

f. **Performance Evaluation**: Optionally, the UI can provide feedback on the performance of the selected machine learning model. This may include metrics such as accuracy, precision, recall, F1-score, and ROC AUC score. Users can use this information to assess the reliability of the fraud detection system and potentially choose a different model if necessary.

g. **User Interaction:** The UI will be designed to be intuitive and user-friendly, allowing users to easily input data, select models, and interpret results. Interactive elements such as dropdown menus, input fields, and buttons can enhance the user experience and make the fraud detection process more efficient.

6. REFERENCE

https://www.kaggle.com/code/janiobachmann/credit-fraud-dealing-with-imbalanced-datasets/notebook

Credit Card Fraud Detection Techniques: A Survey – ScienceOpen

https://www.kaggle.com/code/janiobachmann/credit-fraud-dealing-with-imbalanced-datasets/notebook

https://fraud.net/d/credit-card-fraud-detection/#:~:text=Credit%20card%20fraud%20detection%20employs,to%20recognize%20fraud%2Drelated%20patterns.